

1 SOUND SOURCE LOCALIZATION SYSTEM, AND SOUND REFLECTING ELEMENT

2 **FIELD OF THE INVENTION**

3 The present invention relates to a sound source localization
4 system, a sound source localization method, a sound reflecting
5 element useful for the sound source localization system, and a
6 method for forming the sound reflecting element. It more
7 particularly relates to a high precision sound source
8 localization system, a sound source localization method, a
9 sound reflecting element useful for the sound source
10 localization system, and a method for forming the sound
11 reflecting element, in which the sound source position
12 including the elevation data can be acquired with high
13 precision even if the system comprises a smaller number of
14 microphones.

15 **BACKGROUND OF THE INVENTION**

16 Conventionally, to enhance the sound source localization
17 performance with a microphone array, a processing system
18 capable of making the simultaneous input for multiple channels
19 comprising a number of microphones has been needed. This

1 processing system allows a driving member to be controlled
2 efficiently to face a sound source position. However, if a
3 number of microphones are arranged to acquire the sound source
4 position, there is an inconvenience that the total cost of the
5 system is increased. Therefore, an attempt for reducing the
6 number of microphones has been made. However, in the
7 conventional attempt for reducing the number of microphones, if
8 the number of microphones was reduced, there was an
9 inconvenience that the information for giving a full
10 directionality toward the sound source was lacked. Also,
11 employing the conventional method, there was an inconvenience
12 that the localization of the sound source was more likely to be
13 affected by the surrounding noise, a variation in the property
14 of sound source and the transfer characteristics of the room,
15 although the sound source position was acquired to some extent
16 under the conditions where the properties of the sound source
17 were specified and the measurement environment was managed.

18 In the estimation of the sound source position employing a
19 small number of microphones, various methods have been hitherto
20 proposed. For example, a binaural hearing method employing two
21 microphones has been well known. This method involves using a
22 head transfer function (HRTF), measuring the head transfer
23 function at a binaural position, disposing a sound source for

1 position collected in the scale of angle and range, in addition
2 to the azimuth and elevation.

3 This invention essentially involves a path difference between
4 the sound wave directly collected from the sound source and the
5 reflected wave reflected from a reflecting surface of the sound
6 reflecting element, such that the shape of sound reflecting
7 element is configured to relate the position of sound source
8 with the path difference. In the invention, the sound
9 reflecting element is configured essentially as a set of
10 elliptic curves. Conventionally, for an elliptic curved
11 surface, it is well known that the sound wave produced from one
12 focal point of the ellipse is reflected to the other focal
13 point. Figure 2 shows the typical properties of the ellipse.
14 As shown in Figure 2, the cross section of the reflecting
15 surface is configured using the ellipse in which the sound
16 source is disposed at one focal point A and the microphone is
17 disposed at the other focal point B in this invention. In an
18 arrangement as shown in Figure 2, a sound wave S_r starting from
19 the focal point A is collected at the same focal point B, even
20 if reflected at any position on the wall. Employing the
21 ellipse as the reflecting surface, it follows that the
22 reflected wave always has a certain path difference $(2a-f)$ as
23 defined by the elliptic curve from a sound wave S_d not

1 generating a reference sound at various azimuths, ranges and
2 elevations, and adding the transfer characteristics at the
3 binaural position to acquire the positional information. The
4 above head transfer function is obtained by deciding
5 experimentally the transfer characteristics from the sound
6 source to the ears, including the influences of the head,
7 chest, and concha, for each model, but has a disadvantage of
8 having poor universality.

9 Moreover, the localization of the sound source employing the
10 above head transfer function is made by measuring the signals
11 from the sound source, and selecting a signal consistent with
12 an acoustic spectrum given by the head transfer function
13 measured in advance to acquire the sound source position.
14 Accordingly, the method employing the head transfer function
15 allows the localization of the sound source more or less
16 correctly in principle, if the sound source is a reference
17 sound source. However, since the acquisition of sound source
18 position employing the head transfer function makes the use of
19 a dip or a peak arising in the head transfer function as a
20 characteristic key profile, the sound source position may be
21 possibly misjudged, when the sound source has the dip or peak.
22 Therefore, in the present state of affairs, the acquisition of
23 sound source position employing the head transfer function is

employed more frequently in the sound reproduction than the acquisition of sound source position.

More particularly, the conventional method for acquiring the sound source position was disclosed in Okuno et al., "Are a pair of ears sufficient for robot audition?", The journal of The Acoustical Society of Japan, vol. 58, no. 3, pages 205-210, in 2002, in which the acquisition of sound source position employing two microphones was examined. With this method, the range and azimuth are acquired, employing the ILD (Interaural Level Differences) and the ITD (Interaural Time Difference) obtained from the head transfer function. In the above acquisition of sound source position employing two microphones, the azimuth and range of the sound source can be acquired by measuring the above characteristic values from the acoustic spectrum observed. However, only with these bits of information, the range may not be acquired when the sound source for the acoustic spectrum is located in direct front.

The reason is that in, the interaural level differences and the interaural time difference are constant, even when the range is different. Also, the sound source localization method employing the interaural level differences and the interaural time difference are not effective for vertical localization.

1 The reason is that as long as the azimuth and range are common,
2 the interaural time difference and the interaural level
3 differences are common, even if the elevation varies. From the
4 above reason, to acquire the sound source position including
5 the range and elevation, it is considered that there is a need
6 for taking cues on the reverberation the deformation of the
7 acoustic spectrum, like the monaural hearing as will be
8 described later, and also pointed out that there is a need for
9 further examination.

10 Apart from the binaural hearing, an attempt for acquiring the
11 sound source position by a method of what is called the
12 monaural hearing has been made. The monaural hearing for
13 localization of the sound source is similar to the manner that
14 the man acquires the range to the sound source, in which a
15 larger sound with less reverberation is perceived as the near
16 sound, and a smaller sound with more reverberation is perceived
17 as the distant sound. Employing the loudness of sound and the
18 reverberation as described above, the range to the sound source
19 position is roughly acquired. However, the loudness of sound
20 depends on the sound source of object, and the level of
21 reverberation depends on the experimental environment of
22 acoustic spectrum as well. In the case of man, the information
23 about the sound source of object and the environment, including

1 the visual information, may be compensated by performing a high
2 level information processing, and utilized to acquire the range
3 to the sound source. This processing is practically difficult
4 to implement on a signal processing system comprising an
5 information processing apparatus only based on a pure routine
6 process.

7 Also, in the review for the method for human to acquire the
8 sound source position, it has been found that the azimuth and
9 elevation to the sound source attenuates the spectrum in a
10 specific frequency range under the influence of the head and
11 concha. However, the acquisition method is affected by the
12 properties of the sound source for the same reason as explained
13 for the method employing the head transfer function, and is
14 difficult to implement.

15 Regarding the use of a reflecting plate similar to the concha,
16 a parabolic reflector for collecting a remote subtle sound has
17 been offered by positively utilizing its reflection
18 characteristics. Figure 15 shows a schematic constitution of
19 the parabolic reflector that has been offered. The parabolic
20 reflector 100 as shown in Figure 15 comprises a reflecting
21 plate 102 for reflecting a sound wave 101 from a distant sound
22 source and a microphone 104 for collecting the reflected sound

1 wave. The reflecting plate 102 is roughly formed from a
2 paraboloid, and the microphone 104 is disposed at a focal point
3 position of the paraboloid. The sound wave 106 reflected from
4 the reflecting plate 102 is focused at the focal point to
5 efficiently collect the sound, but there is no function of
6 acquiring the sound source position.

7 Moreover, in an apparatus such as a robot or a sound handling
8 KIOSK terminal that is an object spoken to from the man, it is
9 required to make an operation of "facing in that direction",
10 "turning the directivity of a microphone array to the
11 corresponding direction" or "ignoring a distant sound". For
12 this purpose, it is required that the robot or apparatus
13 recognizes the range or direction to the sound source, or the
14 talker, and controls a drive control system to initiate a
15 necessary operation. That is, under the conditions where the
16 kind of signal sound is unknown, there were the disadvantages
17 with the existing technologies that (1) one microphone does not
18 allow the acquisition of sound source position in principle,
19 and (2) the existing system with two microphones does not allow
20 the acquisition of the range in the forward direction and the
21 elevation in the vertical direction.

22 Also, an increased number of microphones are arranged at

1 appropriate positions as conventionally to relieve the above
2 limitations, whereby the acquisition precision is improved.
3 However, due to a packaging constraint of the design cost, it
4 is sought to relieve the above limitations with a smaller
5 number of microphones.

6 As described above, there is a need for a new method and means
7 suitable for acquiring the position of a sound source,
8 employing an information processing system, without the use of
9 the scale of deformation of spectrum, sound volume or intensity
10 of reverberation needing a high level preliminary knowledge.
11 Also, there is a further demand for a sound source localization
12 system and a sound source localization method in which the
13 range, azimuth and elevation to the sound source are acquired
14 employing the above method and means. Also, there is a further
15 need for a sound reflecting element and a design method for it
16 in which the acquisition of sound source position is
17 excellently made.

18 **SUMMARY OF THE INVENTION**

19 In light of the above-mentioned problems associated with the
20 prior art, an aspect of the present invention recognizes that

1 the disadvantages of the prior art can be solved as far as the
2 elevation information to a sound source can be analyzed with
3 high precision, employing at least one sound collecting means,
4 more particularly, a microphone, whereby a sound source
5 localization system and a sound source localization method are
6 provided with higher precision.

7 In an example embodiment of the present invention, a sound wave
8 generated from a sound source is reflected inherently according
9 to a sound source position, and recorded as the acoustic data
10 collected with the direct sound. This acoustic data is
11 converted into digital data for later processing and once held
12 in a recording unit. The acoustic data can provide a new cue
13 referred to as delay information in this invention. Therefore,
14 in this invention, the new scale of "delay information" is
15 employed in addition to the conventional cue, without depending
16 on the kind of signal sound source, whereby the disadvantages
17 associated with the prior art in the acquisition of sound
18 source position can be solved.

19 In another aspect, to record acoustic data the present
20 invention provides the delay information with a high inherent
21 property, this invention provides a sound reflecting element
22 for reflecting a sound wave generated from the sound source

1 inherently corresponding to a sound source position to enable
2 the recording, and a processing method for processing the
3 recorded acoustic data.

4 In still another aspect, according to the present invention,
5 there is also provided a sound source localization system
6 comprising a sound reflecting element for generating delay
7 information corresponding to a relative position between a
8 sound source and sound collecting means, a storage part for
9 storing the acoustic data collected via the sound reflecting
10 element, and a sound source localization part for acquiring a
11 sound source position, employing the acoustic data on which the
12 delay information is superposed. The sound reflecting element
13 of the invention may be formed as a spheroid associated with
14 the relative position between the sound source and sound
15 collecting means to generate the delay information intrinsic to
16 the relative position. The sound source localization part of
17 the invention may comprise a standard template storage part for
18 storing a standard template containing an intrinsic delay
19 information generated by a white noise sound source, a
20 background noise template storage part for storing a background
21 noise template, a residual generation part for calculating a
22 residual from the acoustic data, employing the standard
23 template and the background noise template, and a selection

1 part for selecting the standard template giving the least
2 residual, employing the generated residual.

3 In another aspect, according to the invention, there is
4 provided a sound source localization method for acquiring the
5 position of a sound source under the control of an information
6 processing apparatus, the method comprising a step of
7 collecting the acoustic data with delay information superposed
8 corresponding to a relative position between a sound source and
9 sound collecting means, a step of storing the collected
10 acoustic data in a storage part, and a step of reading the
11 acoustic data with the delay information superposed and
12 acquiring the relative position of the sound source designated
13 by the delay information.

14 **BRIEF DESCRIPTION OF THE DRAWINGS**

15 The invention and its embodiments will be more fully
16 appreciated by reference to the following detailed description
17 of advantageous and illustrative embodiments in accordance with
18 the present invention when taken in conjunction with the
19 accompanying drawings, in which:

1 Fig. 1 is a view showing the parameters for defining the sound
2 source position and the position in the present invention;

3 Fig. 2 is a view for explaining an essential principle for
4 generating delay information in this invention;

5 Fig. 3 is a view for explaining an essential principle for
6 forming a reflecting surface of a sound reflecting element in
7 this invention;

8 Fig. 4 is a view schematically showing the reflection of sound
9 wave on the reflecting surface as shown in Fig. 3;

10 Fig. 5 is a view showing the envelope for forming the
11 cross-sectional shape of the sound reflecting element formed in
12 this invention;

13 Fig. 6 is a view showing the sound reflecting elements
14 according to an embodiment of the invention;

15 Fig. 7 is a view showing an arrangement of sound reflecting
16 elements according to the embodiment of the invention;

17 Fig. 8 is a schematic flowchart showing a sound source

- 1 localization method of the invention;
- 2 Fig. 9 is a block diagram showing the schematic configuration
3 of a sound source localization system of the invention;
- 4 Fig. 10 is a block diagram showing the detailed configuration
5 of the sound source localization part of the invention;
- 6 Fig. 11 is a view showing a standard template and the storage
7 of three-dimensional position coordinates according to the
8 embodiment of the invention;
- 9 Fig. 12 is a graph showing delay information obtained in this
10 invention;
- 11 Fig. 13 is a graph showing the correlation between the delay
12 information generated in the invention and the delay
13 information on design;
- 14 Fig. 14 is a diagram showing the precision of sound source
15 position acquired in this invention; and
- 16 Fig. 15 is a view showing the schematic configuration of a
17 conventional parabolic reflector.

1 **DESCRIPTION OF SYMBOLS**

2 10 ... sound reflecting element
3 12 ... sound collecting means (microphone)
4 14 ... plane
5 16 ... imaginary line
6 18 ... sound reflecting element
7 20 ... talker
8 22 ... sound reflecting element
9 24 ... recording part
10 26 ... sound source localization part
11 28 ... driving element
12 30 ... acoustic data storage part
13 32 ... STP storage part
14 34 ... BNT storage part
15 36 ... PF part
16 38 ... residual storage part
17 40 ... selection part
18 42 ... application execution part

1 **DETAILED DESCRIPTION OF THE INVENTION**

2 The present invention provides methods, systems and apparatus
3 for solving problems associated with the prior art. The
4 present invention recognizes that the disadvantages of the
5 prior art can be solved as far as the elevation information to
6 a sound source can be analyzed with high precision, employing
7 at least one sound collecting means, more particularly, a
8 microphone, whereby a sound source localization system and a
9 sound source localization method are provided with higher
10 precision.

11 In an example embodiment of the present invention, a sound wave
12 generated from a sound source is reflected inherently according
13 to a sound source position, and recorded as the acoustic data
14 collected with the direct sound. This acoustic data is
15 converted into digital data for later processing and once held
16 in a recording unit. The acoustic data can provide a new cue
17 referred to as delay information in this invention. Therefore,
18 in this invention, the new scale of "delay information" is
19 employed in addition to the conventional cue, without depending
20 on the kind of signal sound source, whereby the disadvantages
21 associated with the prior art in the acquisition of sound

1 source position can be solved.

2 To record the acoustic data by providing the delay information
3 with a high inherent property, this invention provides a sound
4 reflecting element for reflecting a sound wave generated from
5 the sound source inherently corresponding to a sound source
6 position to enable the recording, and a processing method for
7 processing the recorded acoustic data.

8 According to the present invention, there is also provided a
9 sound source localization system comprising a sound reflecting
10 element for generating delay information corresponding to a
11 relative position between a sound source and sound collecting
12 means, a storage part for storing the acoustic data collected
13 via the sound reflecting element, and a sound source
14 localization part for acquiring a sound source position,
15 employing the acoustic data on which the delay information is
16 superposed. The sound reflecting element of the invention may
17 be formed as a spheroid associated with the relative position
18 between the sound source and sound collecting means to generate
19 the delay information intrinsic to the relative position. The
20 sound source localization part of the invention may comprise a
21 standard template storage part for storing a standard template
22 containing an intrinsic delay information generated by a white

1 noise sound source, a background noise template storage part
2 for storing a background noise template, a residual generation
3 part for calculating a residual from the acoustic data,
4 employing the standard template and the background noise
5 template, and a selection part for selecting the standard
6 template giving the least residual, employing the generated
7 residual. The standard template storage part of the invention
8 may store the standard template and the sound source position
9 giving the standard template in association. The sound source
10 localization system of the invention may comprise one or more
11 sound reflecting elements, and simultaneously acquires the
12 positional data of the sound source including a range to the
13 sound source, an azimuth and an elevation as the relative
14 position.

15 According to the invention, there is provided a sound source
16 localization method for acquiring the position of a sound
17 source under the control of an information processing
18 apparatus, the method comprising a step of collecting the
19 acoustic data with delay information superposed corresponding
20 to a relative position between a sound source and sound
21 collecting means, a step of storing the collected acoustic data
22 in a storage part, and a step of reading the acoustic data with
23 the delay information superposed and acquiring the relative

1 position of the sound source designated by the delay
2 information. The delay information of the invention may be
3 generated by reflection from a spheroid associated with the
4 relative position between the sound source and sound collecting
5 means, and the delay information may be generated intrinsic to
6 the relative position. The sound source localization step of
7 the invention may comprise a step of reading out a standard
8 template from a standard template storage part for storing the
9 standard template containing delay information intrinsic to the
10 relative position generated by a white noise sound source, a
11 step of reading out a background noise template from a
12 background noise template storage part for storing the
13 background noise template, a step of calculating a residual
14 from the acoustic data, employing the standard template and the
15 background noise template, and a step of selecting the standard
16 template giving the least residual, employing the generated
17 residual. The selection step of the invention may comprise a
18 step of referring to the selected standard template and
19 acquiring the sound source position corresponding to the
20 standard template. The sound source localization method of
21 this invention may further comprise a step of simultaneously
22 acquiring the range, azimuth and elevation as the relative
23 position from the acquired sound source position to the sound
24 source.

1 According to the invention, there is provided a sound
2 reflecting element for generating delay information
3 corresponding to a relative position between a sound source and
4 sound collecting means, wherein a reflecting surface of the
5 sound reflecting element is designed as an envelope made from a
6 plurality of spheroids that are formed by rotating a plurality
7 of ellipses having the two focal points corresponding to the
8 sound source and the sound collecting mean around an axis
9 connecting the focal points.

10 The plurality of ellipses in this invention may be generated in
11 relation with the elevation between the sound source and the
12 sound collecting means and flatter as the elevation is greater.
13 The reflecting surface in this invention may be designed as an
14 enveloping surface of the plurality of spheroids that are
15 generated by rotating a corresponding ellipse around the axis
16 connecting the focal points.

17 According to the invention, there is provided a formation
18 method of a sound reflecting element for generating delay
19 information corresponding to a relative position between a
20 sound source and sound collecting means, the method comprising
21 a step of generating a plurality of spheroids by rotating an

1 ellipse having the focal points corresponding to the sound :
2 source and the sound collecting mean around an axis connecting
3 the focal points, and a step of forming a reflecting surface by
4 generating an enveloping surface of the plurality of spheroids.
5 The plurality of ellipses in this invention may be generated in
6 relation with the elevation between the sound source and the
7 sound collecting means and flatter as the elevation is greater.

8 A. Constitution of sound reflecting element

9 Figure 1 is a view showing the definition of the range, azimuth :
10 and elevation for use in the present invention. In Figure 1,
11 the microphones M1 and M2 as sound collecting means are
12 employed, in which the azimuth, range and elevation are
13 represented as the position coordinates measured from a middle
14 point between the microphones M1 and M2. A sound source SS is
15 separated away by a predetermined range r from the middle point
16 between the microphones. In the above coordinates, the sound
17 source position can be represented in the Cartesian coordinate
18 system (x, y, z) or polar coordinate system (r, θ, ϕ) in this
19 invention. In the following, the acquisition of elevation is
20 explained as a specific embodiment in this invention, but the :
21 invention is applicable to the acquisition of any sound source

1 reflected and directly going from the focal point A to the
2 focal point B.

3 Taking notice of the path difference, it was reviewed to
4 positively utilize the path difference for the localization of
5 the sound source in this invention. Herein, considering an
6 application mode of the realistic sound reflecting element in
7 the acquisition of sound source position, it is important in
8 the realistic configuration that the microphone is fixed
9 relative to the sound reflecting element, and the sound source
10 such as a talker is moved. Thus, the properties of the
11 reflecting surface are examined, when the position of the
12 microphone is fixed at one focal point B, and the position of
13 the focal point A is changed to have the position of the sound
14 source at the other focal point A. In Figure 3, the maximum
15 range for judging the position of the sound source is defined,
16 and the noise is judged as beyond the maximum range. In Figure
17 3, the sound source is moved from the supposed farthest
18 position f_{\max} to the supposed nearest position f_0 . At the same
19 time, the shape R of an envelope for the ellipses with the
20 focal points f_{\max} and f_0 is shown when the sound source is moved
21 from the supposed farthest position f_{\max} to the supposed nearest
22 position f_0 in Figure 3. As shown in Figure 3, when the focal
23 point A (sound source position) is closer to the microphone,

1 the ellipse has a rounded shape similar to the circle, or when
2 the focal point A (sound source position) is far away from the
3 microphone, the ellipse has a collapsed shape. Also, as the
4 focal point A is farther, the left end shape approximates
5 asymptotically the parabola. In this invention, the shape of
6 sound reflecting element is essentially configured as the
7 envelope of elliptic curves that are formed in connection with
8 the movement of sound source position.

9 Figure 4 is a view schematically showing the reflection of
10 sound wave from the sound source position A, when the
11 reflecting surface is configured as the shape of envelope as
12 shown in Figure 3. As shown in Figure 4, when the sound wave
13 from the nearer sound source position is reflected at a rear
14 portion of the elliptic curve, its reflected wave is collected
15 at the focal point B that is the microphone position. On the
16 other hand, when reflected near an end portion of the elliptic
17 curve, the sound wave is diffused because the angle is not
18 consistent. Therefore, a major portion of the reflected wave
19 detected is occupied by the wave reflected at the rear portion
20 of the sound reflecting element. Similarly, for another sound
21 source position, it has been found that the reflection position
22 to make a major reflected wave component in accordance with its
23 sound source position is generated when the reflecting surface

1 R of sound reflecting element is configured from the envelope.
2 That is, in this invention, it has been found that the major
3 reflected wave intrinsic to the sound source position is
4 generated when the sound reflecting element is formed with the
5 reflecting surface containing the enveloping surface of
6 ellipses. On the other hand, a path difference between the
7 major reflected wave and the direct wave is accompanied with a
8 delay time, which is equivalent to the path difference as
9 defined by the corresponding ellipse.

10 Moreover, the present inventors have reviewed the elevation
11 determination when the envelope of ellipses is employed as the
12 reflecting surface. Figure 5 shows an envelope of elliptic
13 curves and a shape RS of sound reflecting element corresponding
14 to the envelope when the range between the microphone position
15 B and the sound source position A is set at the designed value,
16 and the elevation θ is changed from the supposed lowest angle
17 θ_0 to the supposed highest angle θ_{\max} . As explained in Figure
18 4, if the sound reflecting element RS is formed by the
19 envelope, the sound wave from the sound source at low angle has
20 its major reflected wave reflected at the bottom portion of the
21 sound reflecting element, while the sound wave from the sound
22 source at high angle has its major reflected wave reflected at

1 the top portion of the sound reflecting element. This major
2 reflected wave is accompanied with a delay time corresponding
3 to the path difference defined by the corresponding ellipse. ;
4 That is, the reflected wave intrinsically corresponds to the
5 sound source position.

6 Though this invention has been described above in detail in
7 connection with the cross-sectional shape of the reflecting
8 surface, the shape of the sound reflecting element of the
9 invention is required to be provided in the three dimensions in
10 reality. In this invention, the three-dimensional shape of the
11 reflecting surface of the sound reflecting element for
12 reflecting the sound wave can be formed as the enveloping
13 surface of a plurality of spheroids produced by rotating the ;
14 corresponding ellipse around an axis connecting the focal point
15 on the side where the microphone is placed and the focal point
16 where the sound source position is located.

17 Figure 6 shows a specific embodiment of the sound reflecting
18 element that is configured according to the invention. For the
19 sound reflecting element 10 of the invention as shown in Figure
20 6, the tangential line with each spheroid corresponding to the
21 sound source position is also shown to easily recognize the
22 shape. As shown in Figure 6, the sound reflecting element 10 ;

1 of this invention is configured by cutting the enveloping
2 surface of the spheroid into a size easily employed. Figure 6A
3 is a perspective view of the sound reflecting element 10 as
4 seen from the side of a concave face, and Figure 6B is a
5 perspective view of the sound reflecting element 10 as seen
6 from the side of a convex portion. As shown in Figure 6, the
7 sound reflecting element 10 of the invention has a bottom
8 portion 10a composed of an ellipsoid having a large
9 eccentricity and an upper end portion 10b composed of an
10 ellipsoid having an increased eccentricity, and is narrowed
11 toward the upper end portion 10b in accordance with the
12 elevation.

13 In the sound reflecting element 10 of the invention, the
14 microphone 12 is disposed at one common focal point of the
15 spheroid making up the sound reflecting element 10. Also, the
16 microphone 12 is disposed at a position symmetrical to the
17 sound reflecting element 10 on a plane 14 containing the bottom
18 portion 10a. In the embodiment as shown in Figure 6, the
19 position of the microphone 12 is located on the side of the
20 sound reflecting element 10 above an imaginary line 16
21 connecting the transverse ends of the sound reflecting element
22 10. However, it may take any position as far as the reflected
23 wave from the sound reflecting element 10 is received uniformly

1 with the noise suppressed in this invention. Also, the sound
2 reflecting elements 10 of the invention may be connected
3 vertically with the plane 14 as the boundary.

4 Figure 7 is a perspective view showing an arrangement of the
5 sound reflecting element 10 according to the embodiment of the
6 invention. In the arrangement as shown in Figure 7, the sound
7 reflecting elements 10 and 18 are disposed as one pair. The
8 sound reflecting elements 10 and 18 have the microphones 12 and
9 12a disposed in the same configuration as shown in Figure 6.
10 Moreover, in the arrangement of the sound reflecting element as
11 shown in Figure 7, the sound reflecting elements 10 and 18 are
12 faced in the same direction and suitable for acquiring the
13 sound source position in the direction where the concave
14 portions of the sound reflecting elements 10 and 18 are
15 opposed. The sound reflecting element of the invention can
16 essentially acquire the elevation of the sound source position,
17 employing one sound reflecting element, but employing the sound
18 reflecting elements as one pair as shown in Figure 7, the
19 range, elevation and azimuth to the sound source position may
20 be decided simultaneously.

21 Also, if the overall shape of the sound reflecting element is
22 designed to be small, the path difference between the direct

1 wave and the major reflected wave is shortened. To observe its
2 influence precisely, a high sampling frequency is required. In
3 the specific embodiment of the invention, when the elevation to
4 the sound source is 0° and 72° , and if the path difference
5 between the direct wave and the major reflected wave is about
6 9.5 cm, a delay time difference of about 0.28 ms is produced.
7 When the sampling frequency is 48 KHz, this delay time is
8 equivalent to a difference of about thirteen samples. That is,
9 theoretically, it follows that the elevation to the sound
10 source has a maximum resolution of 13 levels to discriminate
11 the elevation from 0° to 72° . In this invention, if the
12 overall shape is designed to be half in size while keeping the
13 resolution, it is required that the sampling frequency is
14 doubled to 96 KHz. Also, if the overall size of the shape is
15 designed to be double, the same resolution is attained even
16 when the sampling frequency is halved or 24 KHz.

17 B. Sound source localization method and system of the
18 invention

19 Figure 8 is a schematic flowchart of a sound source
20 localization method according to the invention. In the sound
21 source localization method of the invention as shown in Figure
22 9, the acquisition of elevation is made employing the sound

1 reflecting element as explained in the section A. In the sound
2 source localization method of the invention as shown in Figure
3 8, at step S10, the acoustic data such as voice data is
4 collected via the sound reflecting element from the microphone,
5 converted into digital data, employing an AD converter and
6 stored in memory. At step S12, an observed profile is
7 calculated from the acoustic data in accordance with a method
8 as disclosed in detail in "Speech Enhancement by profile
9 fitting method", O. Ichikawa et al., IEICE Transactions on
10 Information and System, VoL. E86-D, No. 3, pp. 514-521, Mar.
11 2003, and at the same time, a standard template (STP) and a
12 background noise template (BNT) that are stored in respective
13 storage parts are read out. At step S14, a residual $\Phi_{n,\omega}$
14 between the observed profile and a linear combination of the
15 standard template and the background noise template is
16 calculated, and stored in an appropriate memory.

17 At step S16, it is determined whether or not there is left any
18 standard template to be further read out. In this manner, the
19 residuals are calculated for all the standard templates. Then,
20 at step S18, the residual $\Phi_{n,\omega}$ is normalized for each subband
21 frequency, and stored in memory. At step S20, the minimum
22 value of the normalized residuals $\Phi_{n,\omega}$ is decided. Then, at

1 step S22, the sound source position corresponding to the
2 standard template giving the minimum value of the calculated
3 residuals is acquired, and selected as the sound source
4 position. At step S24, the coordinates of the sound source
5 position registered corresponding to the selected sound source
6 position are output in an appropriate format to the driving
7 element for controlling the acquired sound source position.

8 As the method for calculating the residual in this invention, a
9 profile fitting method (hereinafter referred to as a PF method)
10 is applied. Particularly in the preferred embodiment of the
11 invention, the PF method is desirably employed. The PF method
12 is a noise suppression method as disclosed in "Speech
13 Enhancement by profile fitting method", O. Ichikawa et al.,
14 IEICE Transactions on Information and System, Vol. E86-D, No.
15 3, pp. 514-521, Mar. 2003, whereby the noise is removed,
16 employing the observed profile from the sound source where the
17 elevation, azimuth and range are defined. However, the PF
18 method is also appropriately employed for a process for
19 estimating the sound source position in this invention.

20 The observed profile for use in a process of the specific
21 embodiment of the invention means a power distribution at each
22 subband frequency that is observed by processing an audio

1 signal recorded by the microphone with a delay sum array, and
2 allocating the angle of directivity of the delay sum array from
3 the maximum value to the minimum value. In this invention, the
4 standard template means a template profile normalized in the
5 area from a two-dimensional observed profile including the
6 delay information recorded via the sound reflecting element
7 employed in the invention and measured in advance for a white
8 noise sound source at the known position in which the direction
9 of allocating the angle of directivity is taken along the axis
10 of abscissas and the power is taken along the axis of
11 ordinates.

12 Also, the background noise template in this invention means a
13 template profile normalized in the area from an acoustic
14 profile observed by placing a white noise sound source at the
15 noise sound source position, in which the width of allocating
16 the angle of directivity is given according to the number of
17 sampling channels. In creating the standard template and the
18 background noise template, it is desirable to employ the white
19 noise having a power over the entire frequency band, as
20 previously described. However, the signal and the noise to be
21 actually observed may be employed to approximate the white
22 noise.

1 Moreover, the residual $\Phi_{n,\omega}$ of the invention is given by the
2 following formula.

3 [Formula 1]

4

5
$$\Phi_{n,\omega} = \int_{\min_{\theta}}^{\max_{\theta}} (X_{\omega}(\theta) - a_{n,\omega} \cdot P_{n,\omega} - \beta_{n,\omega} \cdot Q_{\omega}(\theta))^2 d\theta. \quad (1)$$

6 In the above expression, $X_{\omega}(\theta)$ is the power at the subband
7 frequency ω in which the audio signal with delay information
8 superposed through the sound reflecting element of the
9 invention is processed with the angle of directivity of the
10 delay sum array in the θ direction, and here called the
11 observed profile. $P_{n,\omega}(\theta)$ is the template profile stored as
12 the standard template corresponding to the sound source
13 position, and $Q_{\omega}(\theta)$ is the template profile stored as the
14 background noise template. Also, n corresponds to the sound
15 source position.

16 When the PF method is employed for the sound enhancement, the
17 component decomposition should be made for each frame.

However, for the sound source localization, the component decomposition should be made once for the average over all the audio frames to allow the acquisition of sound source position. So, $X_\omega(\theta)$ may be the average of speaking utterances for several seconds. If $\alpha_{n,\omega}$ and $\beta_{n,\omega}$ are decided using the above formula, the residual $\Phi_{n,\omega}$ is obtained. Moreover, the normalized residual $\bar{\Phi}_{n,\omega}$ is calculated by dividing $\Phi_{n,\omega}$ by the power for each subband and averaging over Ω subbands as defined by the following formula.

[Formula 2]

$$\bar{\Phi}_n = \frac{1}{\Omega} \sum_{\omega} \frac{\Phi_{n,\omega}}{\int_{\min_{\theta}}^{\max_{\theta}} \{X_\omega(\theta)^2 d\theta\}} \quad (2)$$

Also, the acquisition of sound source candidate position is made by selecting a sample template sound source candidate position \hat{n} so that the normalized residual may be the minimized, and selecting the acquired sound source position, using the following formula (3).

[Formula 3]

$$\hat{n} = \arg\min_n (\bar{\Phi}_n) \quad (3)$$

1 An index of "profile" as used in this invention contains not
2 only the cue of delay information for the acoustic spectrum,
3 but also the cues of the interaural time difference and the
4 interaural level differences as conventionally employed. That
5 is, the method of the invention not only detects the delay
6 information, but also makes it possible to employ the cues of
7 the interaural time difference and the interaural level
8 difference as conventionally employed, together with the cue of
9 delay information. Therefore, in this invention, the range,
10 azimuth and elevation required for the acquisition of sound
11 source position can be acquired simultaneously. Accordingly,
12 in the invention, the process for acquiring the sound source
13 position is performed seamlessly, employing a smaller number of
14 microphones than conventionally needed, and the availability of
15 the sound source localization system is expanded. That is, the
16 acquisition of elevation, which was conventionally impossible
17 with the sound source localization method employing as few as
18 one or two microphones, is not dealt with exceptionally, but is
19 processed at the same time with the case of acquiring the angle
20 in the horizontal direction which was conventionally allowed,
21 whereby the process is performed faster. Also, the cue of
22 delay information with the sound reflecting element is added to
23 the case for acquiring the angle which was conventionally

1 allowed, whereby the higher precision localization is allowed.

2 Figure 9 is a view showing the schematic configuration of the
3 sound source localization system according to a specific
4 embodiment of the invention. The sound source localization
5 system of this invention comprises a sound reflecting element
6 22 for collecting and recording voices from the talker 20, a
7 recording part 24 for converting the acoustic data recorded in
8 the sound reflecting element 22 into digital data and storing
9 it, and a sound source localization part 26 for acquiring the
10 sound source position by analyzing the acoustic data. The
11 acquired sound source position information is passed to an
12 application execution part, not shown, in an appropriate format
13 of the coordinates of sound source position such as the
14 Cartesian coordinates (x, y, z) or the polar coordinates $(r, \theta,$
15 $\phi)$ that is decided employing the registered standard template.

16 The application execution part receives an input of position
17 coordinates and drives the driving element 28 needed in the
18 specific embodiment. The driving element 28 may be a head, a
19 hand, a foot, an eye, a mouth, the body, a leg, or the whole
20 body for the robot, a camera or a microphone for the kiosk
21 apparatus, or a microphone or a camera for a security system.

1 However, the invention is not limited to the above driving
2 elements.

3 Also, the sound source localization system of the invention is
4 implemented as an information processing apparatus roughly
5 comprising a CPU (Central Processing Unit), a memory, an
6 external I/O control device, a modem and an NIC. Moreover, the
7 sound source localization system of the invention is mounted on
8 the apparatus comprising the driving element for the robot
9 being driven by application software, in which a predetermined
10 position of the driving element is controlled and driven by
11 comparing a range difference, an azimuth difference and an
12 elevation difference between the original position and the
13 acquired sound source position.

14 Figure 10 is a detailed functional block diagram showing the
15 functional configuration of a sound source localization part 26
16 included in the sound source localization system of the
17 invention. The sound source localization part 26 shown in
18 Figure 10 is realized by a program executing the sound source
19 localization method that is mounted on the robot, kiosk, cache
20 dispenser, a security device for making an operation by sensing
21 a sound, the program being executed by the CPU to function as
22 each means as mentioned above. As shown in Figure 10, the

1 sound source localization part 26 of the invention comprises an
2 acoustic data storage part 30 for reading out the acoustic data
3 once stored in the recording part as the digital data by the
4 sound reflecting element 22, and storing it for processing, a
5 standard template (STP) storage part 32, and a background noise
6 template (BNT) storage part 34.

7 Moreover, the sound source localization part 26 of the
8 invention comprises a profile fitting (PF) part 36 for
9 calculating the residual, a residual storage part 38 for
10 storing the residual $\Phi_{n,\omega}$ obtained by the PF part 36, a
11 selection part 40 for selecting the standard template giving
12 the minimum residual from the normalized residual, and an
13 application execution part 42 for executing a necessary
14 application.

15 The PF part 36 of the invention reads in the acoustic data,
16 converts it into an observed profile, then reads out the
17 standard template from the STP storage part 32, and reads out
18 the background noise template from the BNT storage part 34.
19 The PF part 36 calculates a residual between the linear
20 combination of templates and the observed profile, its result
21 being registered in the residual storage part 38. Moreover,

1 the sound source localization part 26 specifies the normalized
2 residual giving the minimum residual in the selection part 40
3 by normalizing the residual stored in the residual storage part
4 38 and comparing the normalized residuals. Thereafter, the
5 three-dimensional position stored by referring to the standard
6 template giving the corresponding residual is acquired as an
7 appropriate format.

8 Figure 11 is a diagram schematically showing the standard
9 template stored in the STP storage part 32 and the data
10 structure of position coordinates in this invention. The STP
11 storage part 32 is assigned with a memory area corresponding to
12 the three-dimensional position (1, ..., N: N is a positive
13 integer, corresponding to the total number of standard
14 templates). In each memory area i, the STP data and the
15 three-dimensional position data (x, y, z) are stored in
16 association with respective addresses. In another embodiment
17 of the invention, the standard template and the
18 three-dimensional position data may be stored in different
19 memory areas to be referenced from each other.

20 As shown in Figure 11, in the memory area i, the STP data and
21 the three-dimensional position data are stored in association.
22 If the acoustic data is input, the PF part 36 converts it into

1 an observed profile, accesses the memory area i in succession
2 to read out the standard template, calculates the linear
3 combination employing the BNT data, and computes the residual
4 between its value and the observed profile, the result being
5 output to the residual storage part 38. In this invention,
6 delay information defined by the sound reflecting element
7 employed in the invention is introduced into the STP data
8 stored in the STP storage part 32, whereby the elevation is
9 given the intrinsic delay information and acquired with high
10 precision. The selection part 40 refers to the memory area i
11 giving the minimum residual, and reads out the
12 three-dimensional position data (x, y, z) stored in the memory
13 area i to acquire the sound source position. The acquired
14 three-dimensional position data is made a control input into
15 the application execution part 42 to control the driving of the
16 driving element 28, as shown in Figure 11.

17 [Example Embodiments]

18 Specific embodiments of the invention will be described below
19 by way of example, but the invention is not limited to the
20 following examples.

(Example 1)

Sound reflecting element for acquiring the elevation in
the forward direction

Assuming that the azimuth of a sound source candidate position
was 90° (forward direction), the range to a sound source was 2
m, and the acquirable elevation was from 0° to 72° , an
enveloping surface of the spheroid was produced as the sound
reflecting element. An upper end portion of the sound
reflecting element formed in Example 1 reflects a sound wave
from the sound source position at high elevation to converge
into the microphone position and a portion near the root of the
sound reflecting element reflects a sound wave from the sound
source position at low elevation to converge into the
microphone position. On the other hand, the sound wave from
other sound source positions is diffused. If the reflecting
position is different, a stroke difference from the direct wave
is also varied, generating a proper reflected wave with a delay
amount corresponding to the sound source position added.

In the case in which the sound reflecting element was employed,
there was a delay time difference of about 0.28 ms
(milliseconds) in the path difference between the direct wave

1 and the major reflected wave, when the elevation to the sound
2 source was 0° and 72°. The sound source localization system
3 was composed of the sound reflecting element, the microphone,
4 the AD converter, and the microcomputer, whereby the precision
5 of the acquired sound source position was examined. The
6 sampling frequency of the sound source localization system was
7 48 KHz, and the elevation resolution in which the elevation to
8 the sound source was from 0° to 72° was made discernable at 13
9 levels at maximum.

10 (Example 2)

11 Confirmation for generating a "delay information" in
12 the sound reflecting element

13 The sound reflecting elements formed in Example 1 were disposed
14 as shown in Figure 7, and had two microphones attached to form
15 a sound collecting recording part of the invention. For the
16 input, the voices were used, speakings "there" and "hello" for
17 several seconds were regenerated from the sound source position
18 in the forward direction and with the range 2 m and the
19 elevation 0°, 15° 30°, 45° and 60°, whereby an observed profile
20 was produced as the input voice. At this time, the sampling
21 frequency was 48 KHz. To confirm the existence of reflected

1 wave having delay information of the invention, one of the
2 analysis methods of high sensitivity, CSP (Cross-power Spectrum
3 Phase analysis) method by M. Omologo et al. ("Acoustic event
4 localization using a cross power-spectrum phase based
5 technique.", proc. ICASSP 94, pp. 273-276, 1994.) was employed.

6 The CSP method, which traces the acoustic signal at high
7 sensitivity, can give the delay information at high sensitivity,
8 in this invention. For the sound source at an elevation of
9 30°, the calculated CSP coefficients will be shown. Since the
10 CSP method generates a number of pseudo peaks, it is optional
11 how small sub-peak relative to the main peak should be regarded
12 as the valid peak, unlike the main peak. At present, the peaks
13 having one-tenth or more the intensity of the main peak and
14 upper intensities to the third were set as the effective peak.

15

16 Figure 12 shows the CSP coefficients obtained from the input
17 sound signal for the sound source having an elevation of 30°.
18 The results are shown in Table 1.

19

1 [Table 1]

2 Table 1 Peak positions detected by CSP method (unit: number of
3 samples)

Elevation of sound source→	0°	15°	30°	45°	60°
First place peak position	0	0	0	0	0
Second place peak position	N/A	10	9	6	2
Third place peak position	N/A	N/A	N/A	-6	-
Sub-peak position expected on design	±14	±12	±9	±5.5	±2.5

4
5 The peak position having the first place intensity corresponds
6 to the direct wave, in which the peak position 0 indicates that
7 the sound source is disposed in the direct front. At the
8 second place and third place peaks, it is expected that two
9 sub-peaks due to correlation between the direct wave and the
10 reflected wave are detected at the position of designed point
11 as indicated in the table. In Example 2, at least one sub-peak
12 having significant intensity was detected in the cases except
13 for 0° as indicated in the table 1. Also, the delay
14 information for the sound source position was detected by
15 detecting the existence of the expected sub-peak to correspond

1 to the designed point. In the case of the sound source
2 elevation of 0° , the expected sub-peak position was not
3 detected. The reason is that the sound reflecting element
4 formed in Example 1 has a reflection area of zero designed for
5 an elevation of 0° (the root of the sound reflecting element).

6 Figure 13 shows a correlation between the sub-peak position
7 obtained in Example 2, and the sub-peak position expected on
8 design. As shown in Figure 13, the observed sub-peak position
9 has the fine correlation with the existing position of the
10 reflected wave expected in the sound reflecting element of
11 Example 1. From the result of Figure 13, it is found that the
12 sound reflecting element formed in Example 1 gives an expected
13 delay information.

14 (Example 3)

15 Employing the sound reflecting element formed in Example 1, an
16 examination was made to determine whether or not the elevation
17 of sound source could be practically acquired correctly. For
18 the acquisition of sound source position using the delay
19 information, the PF method was employed in this Example 3. A
20 white noise was regenerated from a noise sound source at a

1 horizontal angle 75°, a range 1 m, and an elevation 0° to ;
2 simulate the background noise. The speaking utterances and the
3 sound levels from five positions were produced by changing the
4 elevation, with the background noise superposed, to create the
5 test voices. Employing the following formula, the score was
6 defined from the view point of what difference is provided for
7 the second best candidate, whereby the precision of acquiring
8 the elevation position was examined. Where n^* is an identifier
9 of the standard template corresponding to the correct position,
10 and the residual Φ_{n^*} is the normalized residual at the correct
11 position. ;

12 [Formula 4]

$$13 \quad \rho = \frac{\bar{\Phi}_{\bar{n}} - \bar{\Phi}_{n^*}}{\bar{\Phi}_{\bar{n}}} \quad (4)$$

14 [Formula 5]

$$15 \quad \bar{n} = \underset{n \neq n^*}{\operatorname{argmin}} (\bar{\Phi}_n) \quad (5)$$

16 The above score is given 100% if the normalized residual is
17 zero when the profile corresponding to the correct sound source
18 candidate position is selected, and given 0% or less when the ;
19 acquisition of sound source candidate position fails, because

1 the normalized residual for another profile has the minimum
2 value.

3 In Example 3, the averaging operation of the sub-band when
4 calculating the normalized residual was made in a range from
5 985 Hz to 7504 Hz where the influence of the sound reflecting ;
6 element is most apparent. The results obtained are shown in
7 Figure 14. As shown in Figure 14, in any case, one correct
8 sound source candidate position can be selected from among the
9 five candidate positions by exploiting the component
10 decomposition by the PF method, without being affected by the
11 noise. Also, in this invention, when the background noise
12 template is not employed, the score are decreased with the
13 decrease of the S/N ratio. In this invention, the acquisition
14 of sound source position is made with high precision regardless
15 of the S/N ratio by incorporating the background noise template ;
16 for the residual calculation.

17 Though this invention has been described above by way of
18 example, the invention is not limited to the above described
19 examples. It will be understood to those skilled in the art
20 that various changes and exclusions, and other examples may be
21 made. Also, the sound source acquisition method of the
22 invention can be described in any programming language as ever

1 known, in which these languages include C, C++, Assembler and
2 machine language. Also, the program that can be executed by
3 the computer to perform the sound source acquisition method of
4 the invention may be stored in ROM, EEPROM, flash memory,
5 CD-ROM, DVD, flexible disk, or hard disk and distributed.